

2.4.3. YAPI GEÇERLİLİĞİ:

Beşerî bilimlerde atılganlık, güvensizlik, içe dönüklük vb. özelliklerin ölçülmesi amacıyla ölçme aracı geliştirmek için önce, söz konusu olan özellik ya da özellikleri belirlemek yani yapıyı tanımlamak; daha sonra, tanımlanmış yapıdan sınanabilir denenceler çıkarmak ve çıkarılan denenceleri sınamak için deneysel ve istatistiksel çalışmalar yapmak gerekmektedir. Yapı, birbirleriyle ilgili olduğu düşünülen belli ögelerin ya da ögeler arasındaki ilişkilerin oluşturduğu bir örüntüdür. Bu anlamda, bir testin yapısını geçerleme süreci, temelde testin maddelerine verilen yanıtlar arasındaki ilişkilerin analizine dayanır.

Yapı geçerliliği, bir testin dayandığı kuramsal temelleri ne derece iyi örneklediğiyle ilgilidir. İnsan davranışları ve özellikleri çoğunlukla soyut bir yapıya sahiptir. Zekâ, sevgi, merak, sosyal uyum ve ruh sağlığı gibi soyut kavramları psikolojik testlerle ölçebilmek için önce bu kavramların içeriğinin bilinmesi gerekir. Birey ne tür davranışlar gösterdiği zaman hangi niteliklerin varlığına ya da yokluğuna karar verilebileceğine ilişkin ölçütler geliştirilmesi gerekmektedir.

Yapı geçerliliği bir yandan testin ölçtüğü niteliklerin neler olduğunu araştırma, diğer yandan testi alan kişilerin elde ettikleri puanların ne anlama geldiğini açıklama çabalarıyla ilgilidir. Örneğin bir kişi, geleneksel aile biçimi ile çocuk yetiştirme biçimi arasındaki ilişkiyi incelemek amacıyla bir ölçek geliştirip bu ölçeğin yapı geçerliliğini ortaya koymak istediğinde, geleneksel aile yapısı ve çocuk yetiştirme kavramlarının ne anlamlara geldiğini, ölçme aracındaki maddelerin bu anlamlara uygunluğunu araştırarak yapı geçerliliği konusunda karar verebilir.

2.5. GÜVENİLİRLİK:

Güvenilir bir ölçme aracı, aynı özellikle ilgili olarak arka arkaya yapılan ölçmelerde yaklaşık olarak aynı sayısal sonucu verir; diğer bir ifadeyle bir test, aynı gruba iki ya da üç kez uygulandığında gruptaki her bir kişi bütün uygulamalarda yaklaşık olarak aynı puanı almalıdır. Bir testin ölçmek istediği özelliği ölçebilmesi için o testin söz konusu olan özelliği kararlı olarak ölçmesi gerekir. Güvenirlik çalışmalarının odak noktası şudur: *"Eğer kişi iki defa teste tabi tutulursa iki testten aldığı puanlar birbirine benzer midir ya da birbirine ne kadar yakındır?"* Bu noktada gözlenen puan, gerçek puan, ölçmenin standart hatası ve güvenirlik katsayısı kavramlarına değinmek gerekmektedir.

Örneğin *Murat, üç dakikalık bir sözcük çalışmasında 162 sözcük ya da diğer bir ifadeyle dakikada 54 sözcük yazmıştır. Bu puan Murat'ın becerisini ne ölçüde yansıtmaktadır. Varsayalım ki dakikada 50 sözcük yazılması yeterli kabul edilmektedir. Murat bu düzeyin gerçekten üzerinde midir? Murat geçen hafta dakikada 45 sözcük yazmıştır. Murat'ın bugünkü puanı olanı olan 54, onun kendisini geliştirdiğinin göstergesi midir ya da bir değişim dalgalanması mıdır?*

İki ölçüm arasındaki uyumsuzluğun birçok nedeni olabilir. Bir hareketten diğerine, *"dikkat ve çaba"* değişebilir. Özellikle uzun periyotlarda puan değişmesi, fiziksel büyüme, öğrenme ya da sağlık ve kişilikteki değişimlerden kaynaklanabilir. Yine ikinci ölçümde daha açık soruların kullanılması diğer bir faktör olabilir. İki ölçüm arasındaki puan farklarını yorumlayabilmek için gerçek puan kuramına bakalım.

Psikometride hata kavramı istenmeyen değişkene işaret eder. Ölçme hataları giderilene kadar sürdürülmeli ve böylece gerçek puan elde edilmelidir. Ancak davranış örneği sınırlı olduğu için gözlenen puan gerçek puandan farklılık gösterir. Buradaki farklılık ölçme hatasıdır. Geleneksel olarak hataların varlığı gözlenen puanın gerçek puandan yüksek ya da düşük olmasına neden olur.



Örneğin *yarışlara hazırlanan bir koşucu, bir mesafeyi farklı zamanlarda 23.7, 24.0, 24.2,... 25.1, 25.2 saniyelerde koşmuş olsun. Bu ölçümlerin ortalaması 24.7 ise gerçek puanı 24.7'dir. Koşucu bu puana daha önce 23.7 saniyede koştuğu ölçümü göstererek itiraz edebilir. Bu durumda koşucuya 23.7 ve 25.2 saniyede koştuğu durumlara birçok faktörün etkisinin olabileceği ve 23.7 saniyede tekrar koşmasının belki de hiç olanaklı olamayacağını, gerçek puanının 24.7 olduğunu söyleyebiliriz.* Bu söylem **gerçek puan kuramı**na dayanır.

X = T + E

X = Bireylerin ölçme aracından elde ettiği gözlenen puanı

T = Bireylerin gözlenemeyen gerçek puanı

E = Ölçmeye karışan hata miktarı

Varsayımsal olarak bir ölçme işleminde hata miktarı sıfır ise eşitlik X = T + 0 ve dolayısıyla X = T olmuş olur. Diğer bir ifade ile gözlenen puan, gerçek puana eşit olmuş olur. Ölçme işleminde hata miktarı arttıkça gözlenen puanın gerçek puana olan farkı artmaya başlar. Ölçmenin standart hatası

$$S_h = s\sqrt{1 - r_x}$$

formülü ile bulunur.

S_h= Ölçmenin standart hatası

s= Standart sapma

r_x= Güvenilirlik katsayısı

Bir ölçme işleminde standart hatanın düşük olması, formülden de anlaşılacağı üzere, güvenilirlik katsayısının yüksek, standart sapmanın görece düşük olmasına bağlıdır. Örnek: Bir ölçme işleminde s=8 ve r_x=.75 ise $S_h=8\sqrt{1-.75}=4$. Bu sonuç bireylerin puanlarına –/+ 4 puan hata karıştığı anlamına mı gelir? Gerçek puanın tahmin edilmesinde genellikle üç olasılık değeri kullanılır. Bunlar yaklaşık %68, %95 ve %99'dur. %68 olasılık için bireyin puanından bir standart hata çıkarılır ve puanına eklenirken %95 olasılık için bireyin puanından iki standart hata çıkarılır ve puanına eklenir, son olarak %99 olasılık için bireyin puanından üç standart hata çıkarılır ve puanına eklenir.

%68 olasılık için $X \mp 1 \times S_h$

%95 olasılık için $X \mp 2 \times S_h$

%99 olasılık için $X \mp 3 \times S_h$

Dolayısıyla bir kişi 100 üstünden 50 puan almış ve S_h=4 ise

%68 olasılık: $X \mp 1 \times S_h \Rightarrow 50 \mp 1 \times 4$=Bireyin gerçek puanı %68 olasılıkla 46-54 arasında değişir.

%95 olasılık için $X \mp 2 \times S_h \Rightarrow 50 \mp 2 \times 4$=Bireyin gerçek puanı %95 olasılıkla 42-58 arasında değişir.

%99 olasılık için $X \mp 3 \times S_h \Rightarrow 50 \mp 3 \times 4$=Bireyin gerçek puanı %99 olasılıkla 38-62 arasında değişir.

Ölçmenin standart hatası yükseldikçe ölçme işlemi bireylere rastgele puan vermekten farklı olmayacaktır. Bu nedenle ölçme işlemlerinde güvenilirliğin kestirilmesi çok önemlidir.



Güvenilirlik katsayısı 0 ile 1 arasında değişir. Güvenilirlik katsayısı için alanyazında genellikle 0.70 ve üstü ölçütü kabul edilir. Ancak bu değerin de çok yüksek bir değer olmadığı, 0.70-0.80 arası güvenilirliğin ancak ön bilgi elde etmek amacıyla kullanılabileceği ifade edilir. Genel yetenek gibi bilişsel özellikleri ölçen testlerde güvenilirlik katsayısının 0.90 ve üzerinde olması arzu edilir.



Şekil 8: Güvenilirlik katsayısı

Güvenilirlik, genellikle birden çok uygulamaya dayalı yöntemler ve tek uygulamaya dayalı yöntemler başlıkları altında ele alınır. Birden çok uygulamaya dayalı yöntemler altında test-tekrar test ve eşdeğer (paralel) testler yöntemleri, tek uygulamaya dayalı yöntemler altında ise eşdeğer yarılar, KR-20, KR-21, Cronbach alfa, Hoyt'un varyans analizi, McDonald omega vb. yöntemler yer almaktadır.



Şekil 9: Güvenilirlik kestirim yöntemleri

2.5.1. TEST-TEKRAR TEST YÖNTEMİ: Bu yöntem ile test güvenilirliğini test etmek için bir test, aynı gruba, belli bir zaman aralığıyla iki kez uygulanır. Daha sonra bireylerin birinci uygulamadan aldıkları puanlarla ikinci uygulamadan aldıkları puanlar arasındaki korelasyon hesaplanır. Elde edilen korelasyon katsayısına **kararlılık (devamlılık / istikrarlılık) katsayısı** adı verilir.

	Anksiyete Ölçeği Puanları X_1	Anksiyete Ölçeği Puanları X_2
Ceren	54	50
Filiz	98	110
Aydın	128	120
Ece	89	80
Deniz	101	101
....	87	85
....	98	106
....	67	59
....	122	119
Dursun	79	82

Kararlılık Katsayısı

Korelasyon (r) 1'e yakın

Şekil 10: Test-tekrar test yöntemi çalışma örneği



Güvenilirliğin kararlılık, tutarlılık ve duyarlılık sorunu olduğuna daha önce değinilmişti. Bu yöntemle elde edilen güvenilirlik kanıtı, güvenilirliğin yalnızca kararlılık boyutuna karşılık gelir; güvenilirliğe ilişkin tümel bir sonuç ortaya koymaz. Bu nedenle bir ölçme aracının güvenilirliğine ilişkin kanıt toplamak isteyen bir kişi kararlılığın yanı sıra tutarlılığa ve duyarlılığa ilişkin sorgulamaları yapmak durumundadır.

Bu yöntem, ölçülen özelliğin kararlı olduğu durumlarda uygulanması gereken bir yöntemdir. Testin ölçtüğü özellik sürekli değişkenlik gösteriyorsa bu yöntemle testin güvenirliği hesaplanmamalıdır. Diğer bir deyişle bu yöntem daha çok iki uygulama arasında kolaylıkla değişmeyen özellikleri ölçen testler için uygundur. Örneğin genel zihin yetenekleri, kişilik testleri, ilgi envanterleri, tutum ölçekleri vb. gibi testlerin kararlılık bağlamında güvenilirlikleri bu yöntemle hesaplanabilir.

Bu yolla test güvenilirliğini kestirmede karşılaşılan önemli bir sorun, testin iki uygulanışı arasındaki zaman aralığının ne kadar olması gerektiğidir. Bu soruya kesin bir yanıt vermek olanaksızdır. Aradan geçen zaman, hem ölçme aracının ölçtüğü özellik bakımından yanıtlayıcıların önemli ölçüde değişmelerine hem de birinci uygulamada verilmiş olan yanıtların hatırlanmasına izin vermeyecek bir uzunlukta olmalıdır. Diğer bir ifadeyle iki uygulama arasındaki zamanın belirlenmesindeki temel ilke şudur: "*Birinci uygulamada maddeleri hatırlamayacak kadar uzun, ölçülen özellikte köklü değişimler oluşmayacak kadar kısa olmalıdır.*"

Aradaki zamanın belirlenmesinde dikkat edilmesi gereken bir başka nokta, özelliğin değişim hızıdır. Bazı özellikler daha hızlı değişime açıkken bazıları ise daha zor değişir. Örneğin bireylerin genel yetenek düzeylerinin kısa bir zamanda değişmesi pek olanaklı değilken tutumu görece daha kolay değişebilir. Dolayısıyla bir genel yetenek testi için aradaki zaman daha uzun tutulabilirken tutum ölçeği için daha kısa tutulabilir.

Ayrıca özelliklerin değişimi yaşla da ilişkilidir. Bazı özellikler belli yaşlarda daha durağanken bazı yaşlarda daha değişkendir. Örneğin yaşamın ilk evrelerinde çocukların gelişimi daha hızlı iken yaşın ilerlemesi ile birlikte gelişim yavaşlayabilir. Dolayısıyla bebeklik döneminde uygulanacak bir gelişim envanteri için sürenin daha kısa tutulması gerekebilir.

Test-tekrar test yöntemi, başarı testleri için çok tercih edilmemektedir. Çünkü kısa vadede başarı kolay değişebilen, kararsızlık gösterebilen bir özelliktir. Bu yöntem daha çok yetenek testleri, kişilik envanterleri vb. psikolojik ölçme araçlarından elde edilen puanların güvenilirlik kanıtlarını üretmek için tercih edilmektedir.